

A Simple Alternative to Jet-Clustering Algorithms

Howard Georgi¹

Center for the Fundamental Laws of Nature
Jefferson Physical Laboratory
Harvard University, Cambridge, MA 02138

Abstract

I describe a class of iterative jet algorithms that are based on maximizing a fixed function of the total 4-momentum rather than clustering of pairs of jets. I describe some of the properties of the simplest examples of this class, appropriate for jets at an e^+e^- machine. These examples are sufficiently simple that many features of the jets that they define can be determined analytically with ease. The jets constructed in this way have some potentially useful properties, including a strong form of infrared safety.

Iterative jet clustering algorithms have become an important tool in the analysis of high-energy scattering experiments (see for example [1], and references therein). In this note, I describe a class of iterative jet algorithms that are based on maximizing a fixed function of the total 4-momentum rather than clustering of pairs of jets. I describe some of the properties of the simplest examples of this class. These examples are sufficiently simple that many features of the jets they define can be determined analytically with ease. The jets constructed in this way have some potentially useful properties, including a strong form of infrared safety.

The idea can be stated very simply. Suppose that we have a collection of 4-momenta p_j^μ that we want to organize into jets. In practice, we will typically be interested in masses $\sqrt{p_j^\mu p_{j\mu}}$ that are small compared to their energies and can be ignored in leading order. This is not necessary for the construction, but it leads to considerable simplification, and we will assume that we can set all the particle masses to zero. The jets will then be particular sets of momenta, α with total jet momenta

$$P_\alpha^\mu = \sum_{j \in \alpha} p_j^\mu \quad (1)$$

The underlying assumption is that we want to combine a collection of lines into a single jet, hence increasing the jet energy, if doing so does not increase

¹georgi@physics.harvard.edu

the jet mass to much. So we choose a “jet function” $J(P_\alpha)$ of the total 4-momentum of the ensemble that depends only on the total energy, P_α^0 and the total mass squared divided by energy:

$$J(P) = f(P^0, P^\mu P_\mu / P^0) \quad (2)$$

We require $J(P)$ to have the property that it increases with increasing energy and decreases with increasing m^2/E . We then **find the set α with the maximum value of J** . This gives us our highest J jet (note that it is not necessarily the highest energy jet). Then, as usual in an iterative jet construction, the lines in α are removed and the process is repeated until no lines are left. What makes this different from any pair-wise clustering algorithm that I know of is that we maximize over all possible clusters all at once, rather than building up the jet by clustering pairs. One might worry that this will make the procedure unwieldy for events with many particles. But we will see that at least for one very simple form of the jet function, the clustering is local so that the algorithm can be implemented efficiently. Furthermore the boundaries between the jets have very simple properties that I believe will lead to important simplifications in perturbative calculations (and perhaps beyond).[2]

I should emphasize that in this note, with a jet function like (2) that depends on energy and mass, rather than transverse mass, I am illustrating the idea for jets at an e^+e^- machine, ignoring (for simplicity and because I am not sure how to handle it) the additional complication of hadron beams. I will discuss what I believe is the simplest example of this scheme, in which the function has the form

$$J_\beta(P) = P^0 - \beta P^\mu P_\mu / P^0 \text{ for } \beta > 1 \quad (3)$$

This is monotonically increasing in E and decreasing in m^2 .² As we will see, this produces jets with no fixed “cone size”, but as β increases, there is more of a penalty for large jet mass, and so the jets become more collimated and effectively there is a cone size that decreases as β increases. Obviously, for $\beta = 0$, everything gets included in one “jet”, so this is not particularly interesting. But $\beta > 1$ is interesting and we will be able to understand why analytically.

Let us first consider some general properties of the jet with the largest J_β . Iterating this will give us interesting information about all the jets.

²This is also true for $0 < \beta \leq 1$, but we need $\beta > 1$ for the analysis below.

So we suppose that

$$P_\alpha^\mu \equiv \sum_{j \in \alpha} p_j \quad (4)$$

maximizes J_β . It is obvious that

$$\beta P_\alpha^2 < P_\alpha^{02} \quad (5)$$

We must have $J_\beta(P) > 0$, because there are always positive J_β s (for example single particles) and (5) follows immediately. Then we can write

$$(\beta - 1)P_\alpha^{02} < \beta P_\alpha^2 \quad (6)$$

where $P_\alpha = |\vec{P}_\alpha|$. Hence

$$P_\alpha > \sqrt{\frac{\beta - 1}{\beta}} P_\alpha^0 \quad (7)$$

This shows that for large β , the jets are necessarily nearly light-like.

If there is only one line in the jet, (7) is automatically satisfied for any $\beta > 1$. Suppose that there is more than one line in α and consider a line with 4-momentum p_j^μ for any $j \in \alpha$. Because both p_j^μ and the rest of the 4-momentum of the jet, $P_\alpha^\mu - p_j^\mu$ have lower J_β than P_α^μ by assumption, we can write

$$J_\beta(P_\alpha) > \max(J_\beta(P_\alpha - p_j), J_\beta(p_j)) \quad (8)$$

Let z be the cosine of the angle between \vec{p}_j and the jet direction \vec{P}_α . It is also convenient to define a “jet velocity”

$$v_\alpha \equiv P_\alpha / P_\alpha^0 \quad (9)$$

and the fraction of the jet energy carried by line j

$$r_j \equiv E_j / P_\alpha^0 \quad (10)$$

From (7), we know

$$\sqrt{\frac{\beta - 1}{\beta}} < v_\alpha < 1 \quad (11)$$

and of course from energy conservation,

$$0 < r_j < 1 \quad (12)$$

In terms of v_α and r_j , (8) becomes

$$1 - \beta(1 - v_\alpha^2) > \max\left(1 - r_j - \beta \frac{1 - v_\alpha^2 - 2r_j(1 - z v_\alpha)}{1 - r_j}, r_j\right) \quad (13)$$

This immediately implies a stronger bound on v_α that (7):

$$P_\alpha/P_\alpha^0 = v_\alpha > \sqrt{1 - \frac{1 - r_j}{\beta}} \quad (14)$$

This is not a surprise. It is obvious that if the jet contains a massless particle carrying most of the energy, it is nearly light-like. It is also clear that as $r_j \rightarrow 1$, $z \rightarrow 1$, just by 4-momentum conservation. This means that the bound on jet “size” in the sense of the largest possible angle of a particle in the jet from the jet direction is determined by the soft particles in the jet.

(13) also gives

$$z > \frac{\beta(1 + v_\alpha^2) - (1 - r_j)}{2\beta v_\alpha} \quad (15)$$

or in terms of the angle θ between \vec{p}_j and the jet direction

$$2 \sin \frac{\theta}{2} < \sqrt{\frac{1 - r_j - \beta(1 - v_\alpha)^2}{\beta v_\alpha}} \quad (16)$$

So the maximum angular size of the jet is obtained for soft lines, $r_j \rightarrow 0$, and also depends on the jet velocity. The maximum angular size of the jet is given by

$$2 \arcsin\left(\frac{1}{2\sqrt{\beta}}\right) \text{ as } v_\alpha \rightarrow 1 \quad (17)$$

and it goes to

$$2 \arcsin\left(1 - \sqrt{1 - 1/\beta}\right) \text{ as } v_\alpha \rightarrow \sqrt{1 - 1/\beta} \quad (18)$$

Summarizing the most important result so far, we have found that all the lines in the highest J_β jet are inside a cone of angle

$$\Theta(\beta, v_\alpha) \equiv 2 \arcsin\left(\sqrt{\frac{1 - \beta(1 - v_\alpha)^2}{4\beta v_\alpha}}\right) \quad (19)$$

$$\leq 2 \arcsin\left(1 - \sqrt{1 - 1/\beta}\right) \quad (20)$$

around the jet direction.

We now go on to discuss the particles that are NOT in the highest J_β jet. For a particle with p_j^μ for $j \notin \alpha$, the relations look very similar to (8, 13).

$$J_\beta(P_\alpha) > \max(J_\beta(P_\alpha + p_j), J_\beta(p_j)) \quad (21)$$

or

$$1 - \beta(1 - v_\alpha^2) > \max\left(1 + r_j - \beta \frac{1 - v_\alpha^2 + 2r_j(1 - zv_\alpha)}{1 + r_j}, r_j\right) \quad (22)$$

Note that in this case, while energy conservation does not require $r_j < 1$, (22) does.

(13) also gives

$$z < \frac{\beta(1 + v_\alpha^2) - (1 + r_j)}{2\beta v_\alpha} \quad (23)$$

or in terms of the angle θ between \vec{p}_j and the jet direction

$$2 \sin \frac{\theta}{2} > \sqrt{\frac{1 + r_j - \beta(1 - v_\alpha^2)^2}{\beta v_\alpha}} > 2 \sin \frac{\Theta(\beta, v_\alpha)}{2} \quad (24)$$

This immediately implies that all other particle lines are outside the cone of angular size $\Theta(\beta, v_\alpha)$ around the highest β jet. Furthermore, the particles not in the jet can only approach the jet boundary as $r_j \rightarrow 0$ — that is only for infinitely soft particles.

Thus even though we did not impose a cone size, the jet function J_β did so for us, at least for the highest J_β jet. Particles in the jet are inside the cone. Particles not in the jet are outside the cone. The cone size $\Theta(\beta, v_\alpha)$ varies slightly over the allowed range of v_α , but for large β goes to $1/\sqrt{\beta}$.

Because of this clean separation between the jet defined by J_β and all the other particles, iterating the procedure is absolutely no problem. The jets produced by the procedure are all non-overlapping and bounded by cones of size less than $\Theta(\beta, v_\alpha)$ (each with its own v_α of course). This also insures the IR safety of this procedure. All particles in a particular direction must obviously be in the same jet. In fact, the procedure is VERY IR safe, because particles near the boundary of the cone must get arbitrarily soft.

The relation (18) means that the jets are localized for large β , so one does not have to calculate J_β for all subsets. Instead, the 4π solid angle can be divided into fiducial regions in any convenient way, and the jets in

each fiducial region can be found by looking only at the fiducial region plus the border region within an angle $2 \arcsin\left(1 - \sqrt{1 - 1/\beta}\right)$ of the boundary. This procedure will also give fake jets with direction outside the fiducial region, which should simply be ignored. But for large β , this should enable an enormous improvement in jet-finding efficiency.

I hope I have convinced the reader that the new jet algorithm I am proposing is simple and interesting. It remains to be seen whether it is useful. In closing it is worth emphasizing again that since we have considered a jet function that depends on energy and mass, we are focusing here on a situation like that at an e^+e^- machine where all the jets are on the same footing. Hadron beams and their associated jets introduce an additional complication. It is important to try and generalize this method to hadron colliders because, at least for the moment, that is where all the action is.

Acknowledgements

I am very grateful to Matthew Schwartz and Ilye Feige for many informative conversations about jets, to Steve Ellis for a reality check and to Jakub Scholtz for catching a minor misstatement in the first version, and to Jesse Thaler for helpful comments. I am also extremely grateful to the **Invisibles**, including Belen Gavela and Pilar Hernandez whose invitation to the Invisibles Summer School and Workshop was an important motivation for this work. This work grew initially out of a three-year-long effort with Greg Kestin and Aqil Sajjad to construct an effective field theory on the light shell. I am grateful for their help during this ongoing process. This research has been supported in part by the National Science Foundation under grants PHY-1067976 and PHY-1418114.

References

- [1] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, “Towards an understanding of jet substructure,” *JHEP* **1309** (2013) 029, [arXiv:1307.0007 \[hep-ph\]](#).
- [2] R. Kelley, J. R. Walsh, and S. Zuberi, “Disentangling Clustering Effects in Jet Algorithms,” [arXiv:1203.2923 \[hep-ph\]](#).